Systematic review and critical appraisal of child abuse measurement instruments

Suriati Mohamed Saini , Cassandra R. Hoffmann ,
Christos Pantelis , Ian P. Everall , Chad A. Bousman

PII: S0165-1781(18)31391-X
DOI: https://doi.org/10.1016/j.psychres.2018.12.068
Reference: PSY 11999

To appear in: *Psychiatry Research*

Received date: 2 August 2018
Revised date: 8 November 2018
Accepted date: 11 December 2018

Please cite this article as: Suriati Mohamed Saini , Cassandra R. Hoffmann , Christos Pantelis , Ian P. Everall , Chad A. Bousman , Systematic review and critical appraisal of child abuse measurement instruments, *Psychiatry Research* (2018), doi: https://doi.org/10.1016/j.psychres.2018.12.068

Highlights

- 15% of the child abuse instruments had strong to moderate levels of evidence.
- No instruments had adequate levels of evidence for all COSMIN criteria.
- No single instrument is superior to all others across settings and populations.
- Measures that capture the effects child abuse on brain development are limited.

Systematic review and critical appraisal of child abuse measurement instruments

Suriati Mohamed Saini[a,b], Cassandra R. Hoffmann[a], Christos Pantelis[a,c,d,e], Ian P. Everall[a,c,d,g], Chad A. Bousman[a, f,h,i]

[a] Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health; Parkville, VIC, Australia
[b] Department of Psychiatry, UKM Medical Centre, Jalan Yaacob Latif, Cheras, Kuala Lumpur, Malaysia
[c] Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia
[d] Centre for Neural Engineering, The University of Melbourne, Carlton, VIC, Australia
[e] North Western Mental Health, Melbourne Health, Parkville, VIC, Australia
[f] Alberta Children's Hospital Research Institute, Calgary, Canada
[g]Institute of Psychiatry, Psychology & Neuroscience, King's College London, United Kingdom

[h] Department of Medical Genetics, Psychiatry, and Physiology and Pharmacology, The University of Calgary, Canada

[i] Hotchkiss Brain Institute, University of Calgary, Calgary, Canada

Correspondence:

Dr Chad Bousman, 3330 Hospital Drive NW, Alberta Children's Hospital Research Institute, University of Calgary, AB, T2N 4N1, Canada. chad.bousman@ucalgary.ca,
T: +1 403.441.8409

Abstract

Child abuse is a major public health concern and a strong predictor of adult psychopathology. However, a consensus on how best to measure child abuse is not evident. This review aimed to critically appraise the methodological quality and measurement properties of published child abuse measures, examined the strength of evidence of these instruments for research use using the COnsensus-based Standards for the selection of health Measurement InstrumeNts (COSMIN) checklist and determined which measures were capable of providing information on the developmental timing of abuse. Systematic search of electronic databases identified 52 eligible instruments from 2095 studies. Only 15% (n=8) of the instruments had strong to moderate levels of evidence for three or more of the nine COSMIN criteria. No instrument had

adequate levels of evidence for all criteria, and no criteria were met by all instruments. Our results indicate there is no single instrument that is superior to all others across settings and populations. The availability of measures capable of capturing the effects of child abuse on brain development and associated behavioral phenotypes are limited. Refined instruments with a focus on capturing abuse events during development are warranted in addition to further evaluation of the psychometric properties of these instruments.

Keywords: child, abuse, maltreatment; measure; instrument, brain development; sensitive period

1. Introduction

Child abuse is a major public health concern. A recent review of a series of meta-analyses reported that prevalence rates of child abuse for self-report studies are 12.7% for sexual abuse, 22.6% for physical abuse, 6.3% for emotional abuse, 16.3% for physical neglect and 18.4% for emotional neglect (Stoltenborgh et al., 2015). Individuals who experience abuse during childhood are more likely to develop depression (Infurna et al., 2016), post-traumatic stress disorder (Breslau et al., 2014; Brewin et al., 2000), schizophrenia (McGrath et al., 2017; Varese et al., 2012), dissociative disorder (Draijer and Langeland, 1999), personality disorder (Johnson et al., 1999), substance use (Evans et al., 2017; Hamburger et al., 2008; Lo and Cheng, 2007), and suicidal behaviour (Dube et al., 2001; Stansfeld et al., 2017; Tunnard et al., 2014). The total lifetime estimated economic cost of child abuse is estimated at USD124 billion in the United States each year (Fang et al., 2012).

Despite the establishment of this potent and robust marker of psychiatric risk and proliferation of childhood abuse instruments, scientific gaps remain. First, although several reviews have evaluated child abuse measurement instruments (Burgermeister, 2007; Hulme, 2004; Roy and Perry, 2004; Satapathy et al., 2017; Strand et al., 2005), consensus on which child abuse instruments provide the best methodological quality, and measurement properties has not been achieved. Previous reviews did not conduct a systematic evaluation of the methodological quality and measurement properties of available child abuse instruments using the COnsensus-based Standards for the selection of health Measurement InstrumeNts (COSMIN) checklist (Mokkink et al., 2010). Thus, the strength of evidence for each instrument based on methodological quality, and measurement properties are largely unknown. The COSMIN initiative aims to facilitate the selection of high-quality patient-reported outcome measures for research and clinical practice and provide a comprehensive guideline for systematic review of the patient-reported outcome measures (Prinsen et al., 2018).

Second, less attention has been given to the development of psychometrically sound measures that allow for the assessment of time-dependent effects of child abuse across development. Emerging evidence has provided insight into the plasticity of the developing brain as a function of experience and has been a focus of research in recent years. Abuse during sensitive periods (e.g.: prenatal, postnatal and pubertal periods) is known to exert maximal effects on the developmental trajectory of specific brain regions (Andersen, 2003; Bale, 2015; Choi et al., 2012; Curley and Champagne, 2016; Pechtel et al., 2014; Tomoda et al., 2009; Tomoda et al., 2012). Thus, information on the timing of abuse exposure is essential in facilitating a more comprehensive description of the long-term effects of child abuse.

As such, we systematically reviewed, critically appraised, compared, and summarized the methodological quality and measurement properties of published child abuse measures using the COSMIN checklist. In addition, we assessed the strength of evidence of these measurement instruments for research use and determined which measures were capable of providing information on the developmental timing of child abuse.

2. Materials and method

2.1. Defining child abuse

The WHO Consultation on Child Abuse Prevention states: 'Child abuse or maltreatment constitutes all forms of physical and/or emotional ill-treatment, sexual abuse or negligent treatment or commercial or other exploitation, resulting in actual or potential harm to the child's health, survival, development or dignity in the context of a relationship of responsibility, trust or power' (World Health Organization, 1999). In this review we focussed on instruments that measure one or more of the five main subtypes of child abuse by caregivers, including physical abuse (acts of commission that cause actual physical harm or have the potential for harm), sexual abuse (acts where a caregiver uses a child for a sexual gratification), emotional abuse (failure of a caregiver to provide appropriate and supportive environments) and neglect (failure of a parent to provide for physical and emotional development of child) (Daro and Dodge, 2009; Mersky et al., 2017; Runyan et al., 2002).

2.2 Search strategy and data extraction

A systematic search of three electronic databases: Ovid (Medline), PsycINFO and Health and Psychosocial Instruments was performed using the following keywords: 'child or

childhood or early life' and 'abuse or maltreatment or trauma or abuse or adversity' and 'instrument or questionnaire or tool.' The search was restricted to articles in English, and those published until July 2016. Only instruments that measured child abuse were included. Instruments for screening trauma-related disorder or symptoms (e.g., post-traumatic stress disorder, dissociative disorder, depression) were not evaluated. Bibliographies of selected articles were also screened to identify other relevant instruments. When neccesary, attempts were made to contact corresponding authors of the included studies to obtain missing information during the data extraction process.

Two independent reviewers (SMS and CRH) screened the titles and abstracts of all articles identified by the search strategies and assessed the full-text copies of the relevant articles. Among eligible studies the following data were extracted: i) characteristics of included studies such as instrument names and authors, age specifier, dimension(s) measured, response format, time to administer, population assessed, country where the study was conducted and psychometric properties; ii) evaluation of measurement properties; iii) interpretability and generalizability of the results. The 'preferred reporting items for systematic review and meta-analysis protocols' (PRISMA-P) (Shamseer et al., 2015) was followed in the reporting of this systematic review.

2.3. Methodological and measurement quality assessment

Assessment of the methodological quality of the instruments was done based on the COSMIN checklist (Mokkink et al., 2010). The checklist consists of 114 items, across twelve criteria. Nine of the criteria pertain to internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross-cultural validity, criterion validity, and responsiveness. The remaining three criteria pertain to item response theory methods (when applicable), interpretability, and generalizability of the instrument properties.

Each of the 114 items is scored on a 4-point rating scale (excellent, good, fair and poor) and within each of the twelve criteria, the item with the lowest rating is assigned to that criterion (De Vet et al., 2011; Terwee et al., 2012).

Measurement quality ratings for each instrument were determined according to the COSMIN taxonomy, and rated as 'adequate' (+), 'not adequate' (-) or 'unclear' (?), based on the predefined criteria (De Vet et al., 2011; Mokkink et al., 2010) (Table 1).

2.4. The strength of evidence assessment

The strength of evidence for each instrument was categorized as strong, moderate, limited, conflicting, or unknown based on the methodological and measurement quality as well as the number and consistency of results between psychometric studies of each instrument. S*trong* evidence included several methodologically good articles or one excellent article, reporting consistent evidence for or against measurement properties. *Moderate* evidence indicated the presence of several methodologically fair, or one good study. Whereas, *limited* evidence was assigned to instruments with one study of fair quality and *conflicting* evidence was assigned to instruments with conflicting/mixed findings. *Unknown* evidence indicated the presence of studies of 'poor' methodological quality or the absence of studies (De Vet et al., 2011)

3. Results

A total of 2095 articles were identified, screened, and assessed for eligibility resulting in 68 studies representing 52 child abuse measurement instruments (Figure 1).

3.1. Characteristics of child abuse measures

Characteristics of the included measures are summarized in Appendix Table A.1. Most instruments were self-report, retrospectively measured child abuse before the age of 18,

and could be administered in less than 30 minutes. Instrument development and psychometric evaluation primarily utilized convenient female samples (undergraduates, outpatients) ranging from 17 to 1,978 participants. Most measures were developed in English, although some had been adapted and translated into other languages. All instruments were evaluated using classical test theory except for the Maltreatment and Abuse Chronology of Exposure (MACE) scale which was developed using item response theory.

3.2. Methodological and measurement quality

Methodological and measurement quality of the eligible studies are summarized for each criterion in Table 2. None of the 52 instruments identified could be assessed for all criteria based on the literature. In fact, most instruments did not have information related to their internal consistency (58%), reliability (56%), measurement error (98%), content validity (62%), structural validity (90%), hypothesis testing (56%), cross-cultural validity (94%), criterion validity (85%), or responsiveness (100%).

3.3. Strength of evidence

The strength of evidence for each instrument is summarized in Figure 2. No instrument had strong to moderate strength of evidence for all nine COSMIN criteria, and none of the criteria was met by all instruments. The strongest evidence was available for the CTQ with strong to moderate evidence for 55% of the COSMIN criteria, followed by the CTQ-SF and MACE, which both had strong to moderate evidence for 44% of the criteria. At the criterion level, a quarter or more of the assessed instruments had strong to moderate evidence for content validity (34.6%), reliability (46.2%), and internal consistency (26.9%).

None of the instruments assessed had a strong strength of evidence for criterion related to measurement error, cross-cultural validity, criterion validity, or responsiveness.

4. Discussion

We found a wide variation in methodology, measurement, and psychometric properties among instruments used to measure child abuse. Our findings suggest there is no single instrument that is superior to all others across settings and populations given that several criteria in the COSMIN checklist such as measurement error, criterion validity and cross-cultural validity were not evaluated or reported by most of the instruments we assessed. As such the most appropriate instrument will depend largely on the context in which it will be employed. However, the findings from this systematic review did identify a number of instruments worthy of particular attention.

Of the 52 measures, CTQ is the only scale that has been thoroughly investigated and demonstrates a strong level of evidence with adequate internal consistency, reliability, content validity, structural validity and convergent (hypothesis testing) validity. It is also widely used and has been translated into many languages. Good alternatives to the CTQ included CTQ-SF, MACE, CAT, ETI-SR, AEIII, CCMI, PMI, ADQ, BCAS, ETI and CEVQ which demonstrated moderate to strong evidence for validity and reliability. Thus, these measures would be preferred over the others we assessed, assuming additional factors (e.g., setting and population appropriateness, alignment with study hypotheses, availability) related to instrument selection are equal. However, we cannot fully endorse these instruments given that several other criteria in the COSMIN checklist were not evaluated or reported.

Interpretability encompasses a number of important characteristics of an instrument such as generalizability and floor or ceiling effects (De Vet et al., 2011). None of the studies

mentioned procedures for handling missing values or how missing values could affect the generalizability of the instrument. Furthermore, only six scales examined the distribution of scores in the study samples. Distribution of the scores produced by an instrument is considered an important characteristic in that it provides a proper interpretation of the scores on a measurement instrument and a proper interpretation of the measurement properties (De Vet et al., 2011; Terwee et al., 2012). Finally, there is a lack of investigations assessing measurement invariance for cross-cultural validity studies using confirmatory factor analysis or differential item functioning. Thus, the applicability of most of the instruments beyond Caucasians in Western cultures is limited and requires more research.

Beyond methodological and measurement quality, we were also interested in determining which measures were capable of providing information on the developmental timing of child abuse. Eight of the reviewed measures (TAI, SLEI, SPAQ, ETI, ETI-SR, CCMI, LTVH, and MACE) have been shown to delineate the developmental stages. However, only five (ETI, ETI-SR, CCMI, LTVH, MACE) were shown to have moderate to strong level of evidence in this review. The ETI, ETI-SR, and CCMI examine the frequency of child abuse by developmental stage and relationship with the perpetrator (Bremner et al., 2007; Riddle and Aponte, 1999). The LTVH enquiries about trauma and victimization experiences, frequency, duration, and degree of danger and fear experienced using a semi-structured questionnaire (Widom et al., 2005). Measures that retrospectively assess child abuse have limitations in capturing detailed information on how exposure levels changed across development and delineated sensitive period effects for developmental studies (Teicher and Parigger, 2015). The MACE is a new retrospective instrument that primarily developed to overcome these limitations by assessing cumulative severity and number of types (multiplicity) of recollected exposure to abuse during each year of childhood (Teicher and Parigger, 2015). The MACE also provides features to solicit additional information on

peer victimization, witnessing inter-parental physical violence and violence towards siblings and eliminates items that could confound exposure with familial risk (such as parental loss, unavailability of father or unavailability of a mother for enumerated good reasons) (Teicher and Parigger, 2015).

There are some caveats in this systematic review that should be noted. First, only studies published in English were considered, resulting in the exclusion of four measures that were developed, translated and published in other languages (Aslan and Alparslan, 1999; Wingenfeld et al., 2010; Yu et al., 2009; Zhao et al., 2005). Second, although the COSMIN checklist is validated and well structured, there are still elements of subjectivity due to the different methodologies employed and information reported by the authors that might affect the decision in the rating process. Third, there has been little credence given to the reliability and validity of retrospective measures of child abuse. Autobiographical memory, memory impairment due to psychopathology, and/or mood-congruent memory retrieval biases are potential threats to the reliability and validity of self-reported retrospective instruments of child abuse (Brewin et al., 1993). However, the credibility of this potential threat was not supported by a comprehensive review that suggested high stability in the recall, even with changes in mood and clinical status (Brewin et al., 1993). The accuracy of memories for early experiences depends largely on the characteristics of the events or events to be recalled (Brewin et al., 1993). Furthermore, a more recent study comparing the identification of childhood maltreatment using prospective case review and retrospective self-report methodologies found that both methods identified severe cases of maltreatment, but neither method on its own captured all cases of childhood maltreatment (Shaffer et al., 2008).

5. Conclusion

Overall findings from this systematic review and critical appraisal of child abuse measurement instruments suggest no single instrument is superior to all others across settings and populations. There are a number of instruments with moderate to strong level of evidence for methodological quality and measurement properties that may be suitable for particular research questions. The present review also revealed that the availability of measures capable of capturing the effects of child abuse on brain development and associated behavioral phenotypes is severely limited. Refined instruments with a focus on capturing adverse events during sensitive periods of development are warranted in addition to further evaluation of the psychometric properties of these instruments. Without these future efforts, the ability to detect reliable and valid effects of child abuse on unfavourable brain and related behavioral outcomes will be hindered, and potential interventions to prevent these detrimental outcomes will be delayed.

Acknowledgements

None of the funding sources played any role in the study design, analysis or interpretation of data, in the writing of the report, or in the decision to submit the paper for publication. SMS was supported by a Ministry of Higher Education Malaysia Scholarship and the National University of Malaysia. CP was supported by a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellowship (ID: 1105825), and a Brain and Behavior Research Foundation (NARSAD) Distinguished Investigator Award. CAB was supported by the Alberta Children's Hospital Research Foundation and the University of Calgary Cumming School of Medicine.

Conflict of interest

No conflicts of interest to declare.

References

Andersen, S.L., 2003. Trajectories of brain development: point of vulnerability or window of opportunity? Neuroscience and Biobehavioral Reviews 27 (1), 3-18.

Aslan, S.H., Alparslan, Z.N., 1999. The reliability, validity and factor structure of the Childhood Trauma Questionnaire among a group of university students. Turkish Journal of Psychiatry 10 (4), 275-285.

Bale, T.L., 2015. Epigenetic and transgenerational reprogramming of brain development. Nature Reviews Neuroscience 16 (6), 332-344.

Bremner, J.D., Bolus, R., Mayer, E.A., 2007. Psychometric properties of the early trauma inventory–self report. The Journal of nervous and mental disease 195 (3), 211-218.

Breslau, N., Koenen, K.C., Luo, Z., Agnew-Blais, J., Swanson, S., Houts, R.M., Poulton, R., Moffitt, T.E., 2014. Childhood maltreatment, juvenile disorders and adult post-traumatic stress disorder: a prospective investigation. Psychological Medicine 44 (9), 1937-1945.

Brewin, C.R., Andrews, B., Gotlib, I.H., 1993. Psychopathology and early experience: a reappraisal of retrospective reports. Psychological Bulletin 113 (1), 82-98.

Brewin, C.R., Andrews, B., Valentine, J.D., 2000. Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. Journal of Consulting and Clinical Psychology 68, 748-766.

Burgermeister, D., 2007. Childhood adversity: a review of measurement instruments. Journal of Nursing Measurement 15 (3), 163-176.

Choi, J., Jeong, B., Polcari, A., Rohan, M.L., Teicher, M.H., 2012. Reduced fractional anisotropy in the visual limbic pathway of young adults witnessing domestic violence in childhood. Neuroimage 59 (2), 1071-1079.

Curley, J.P., Champagne, F.A., 2016. Influence of maternal care on the developing brain: Mechanisms, temporal dynamics and sensitive periods. Frontiers in Neuroendocrinology 40, 52-66.

Daro, D., Dodge, K.A., 2009. Creating community responsibility for child protection: Possibilities and challenges. The Future of Children/Center for the Future of Children, the David and Lucile Packard Foundation 19 (2), 67.

De Vet, H.C., Terwee, C.B., Mokkink, L.B., Knol, D.L., 2011. Measurement in medicine: a practical guide. Cambridge University Press.

Draijer, N., Langeland, W., 1999. Childhood trauma and perceived parental dysfunction in the etiology of dissociative symptoms in psychiatric inpatients. American Journal of Psychiatry 156 (3), 379-385.

Dube, S.R., Anda, R.F., Felitti, V.J., Chapman, D.P., Williamson, D.F., Giles, W.H., 2001. Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the Adverse Childhood Experiences Study. JAMA 286 (24), 3089-3096.

Evans, E., Upchurch, D., Grella, C.E., 2017. Gender differences in the effect of childhood adversity on alcohol, drug, and poly-substance use disorders. Drug and Alcohol Dependence 171, e60.

Fang, X., Brown, D.S., Florence, C.S., Mercy, J.A., 2012. The economic burden of child maltreatment in the United States and implications for prevention. Child Abuse and Neglect 36 (2), 156-165.

Hamburger, M.E., Leeb, R.T., Swahn, M.H., 2008. Childhood maltreatment and early alcohol use among high-risk adolescents. Journal of Studies on Alcohol and Drugs 69 (2), 291-295.

Hulme, P.A., 2004. Retrospective measurement of childhood sexual abuse: A review of instruments. Child maltreatment 9 (2), 201-217.

Infurna, M.R., Reichl, C., Parzer, P., Schimmenti, A., Bifulco, A., Kaess, M., 2016. Associations between depression and specific childhood experiences of abuse and neglect: a meta-analysis. Journal of Affective Disorders 190, 47-55.

Johnson, J.G., Cohen, P., Brown, J., Smailes, E.M., Bernstein, D.P., 1999. Childhood maltreatment increases risk for personality disorders during early adulthood. Archives of General Psychiatry 56 (7), 600-606.

Lo, C.C., Cheng, T.C., 2007. The impact of childhood maltreatment on young adults' substance abuse. The American journal of drug and alcohol abuse 33 (1), 139-146.

McGrath, J., McLaughlin, K., Saha, S., Aguilar-Gaxiola, S., Al-Hamzawi, A., Alonso, J., Bruffaerts, R., de Girolamo, G., de Jonge, P., Esan, O., 2017. The association between childhood adversities and subsequent first onset of psychotic experiences: a cross-national analysis of 23 998 respondents from 17 countries. Psychological Medicine 47 (7), 1230-1245.

Mersky, J.P., Janczewski, C.E., Topitzes, J., 2017. Rethinking the measurement of adversity: moving toward second-generation research on adverse childhood experiences. Child maltreatment 22 (1), 58-68.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., De Vet, H.C., 2010. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research 19 (4), 539-549.

Pechtel, P., Lyons-Ruth, K., Anderson, C.M., Teicher, M.H., 2014. Sensitive periods of amygdala development: the role of maltreatment in preadolescence. Neuroimage 97, 236-244.

Prinsen, C., Mokkink, L., Bouter, L., Alonso, J., Patrick, D., de Vet, H., Terwee, C., 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research 27 (5), 1147-1157.

Riddle, K.P., Aponte, J.F., 1999. The comprehensive childhood maltreatment inventory: Early development and reliability analyses. Child Abuse and Neglect 23 (11), 1103-1115.

Roy, C.A., Perry, J.C., 2004. Instruments for the assessment of childhood trauma in adults. The Journal of nervous and mental disease 192 (5), 343-351.

Runyan, D., Wattam, C., Ikeda, R., Hassan, F., Ramiro, L., 2002. Child abuse and neglect by parents and other caregivers. In: Krug EG Dahlberg LL, Mercy JA, Zwi AB, Lozano R, editors. World report on violence and health. Geneva: World Health Organization. .

Satapathy, S., Choudhary, V., Sagar, R., 2017. Tools to assess psychological trauma & its correlates in child sexual abuse: A review & current needs in Asia. Asian Journal of Psychiatry 25, 60-73.

Shaffer, A., Huston, L., Egeland, B., 2008. Identification of child maltreatment using prospective and self-report methodologies: A comparison of maltreatment incidence and relation to later psychopathology. Child Abuse and Neglect 32 (7), 682-692.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 4 (1), 1-9.

Stansfeld, S.A., Clark, C., Smuk, M., Power, C., Davidson, T., Rodgers, B., 2017. Childhood adversity and midlife suicidal ideation. Psychological Medicine 47 (2), 327-340.

Stoltenborgh, M., Bakermans-Kranenburg, M.J., Alink, L.R., van Ijzendoorn, M.H., 2015. The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. Child Abuse Review 24 (1), 37-50.

Strand, V.C., Pasquale, L.E., Sarmiento, T.L., 2005. Child and adolescent trauma measures: A review. Trauma Violence Abuse 6 (1), 55-78.

Teicher, M.H., Parigger, A., 2015. The 'Maltreatment and Abuse Chronology of Exposure'(MACE) scale for the retrospective assessment of abuse and neglect during development. PloS One 10 (2), e0117423.

Terwee, C.B., Mokkink, L.B., Knol, D.L., Ostelo, R.W., Bouter, L.M., de Vet, H.C., 2012. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Quality of Life Research 21 (4), 651-657.

Tomoda, A., Navalta, C.P., Polcari, A., Sadato, N., Teicher, M.H., 2009. Childhood sexual abuse is associated with reduced gray matter volume in visual cortex of young women. Biological Psychiatry 66 (7), 642-648.

Tomoda, A., Polcari, A., Anderson, C.M., Teicher, M.H., 2012. Reduced visual cortex gray matter volume and thickness in young adults who witnessed domestic violence during childhood. PloS One 7 (12), e52528.

Tunnard, C., Rane, L.J., Wooderson, S.C., Markopoulou, K., Poon, L., Fekadu, A., Juruena, M., Cleare, A.J., 2014. The impact of childhood adversity on suicidality and clinical course in treatment-resistant depression. Journal of Affective Disorders 152, 122-130.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., Read, J., van Os, J., Bentall, R.P., 2012. Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. Schizophrenia Bulletin 38 (4), 661-671.

Widom, C.S., Dutton, M.A., Czaja, S.J., DuMont, K.A., 2005. Development and validation of a new instrument to assess lifetime trauma and victimization history. Journal of Traumatic Stress 18 (5), 519-532.

Wingenfeld, K., Spitzer, C., Mensebach, C., Grabe, H.J., Hill, A., Gast, U., Schlosser, N., Höpp, H., Beblo, T., Driessen, M., 2010. The German version of the Childhood Trauma Questionnaire (CTQ): preliminary psychometric properties. Psychotherapie, Psychosomatik, Medizinische Psychologie 60 (11), 442-450.

World Health Organization, 1999. Report of the Consultation on Child Abuse Prevention, 29-31 March 1999, WHO, Geneva.

Yu, J., Park, J., Park, D., Ryu, S., Ha, J., 2009. Validation of the Korean Childhood Trauma Questionnaire: the practical use in counselling and therapeutic intervention. Korean J Health Psychol 14 (3), 563-578.

Zhao, X., Zhang, Y., Li, L., Zhou, Y., Li, H., Yang, S., 2005. Reliability and validity of the Chinese version of Childhood Trauma Questionnaire. Chinese Journal of Clinical Rehabilitation 9 (20), 105-107.

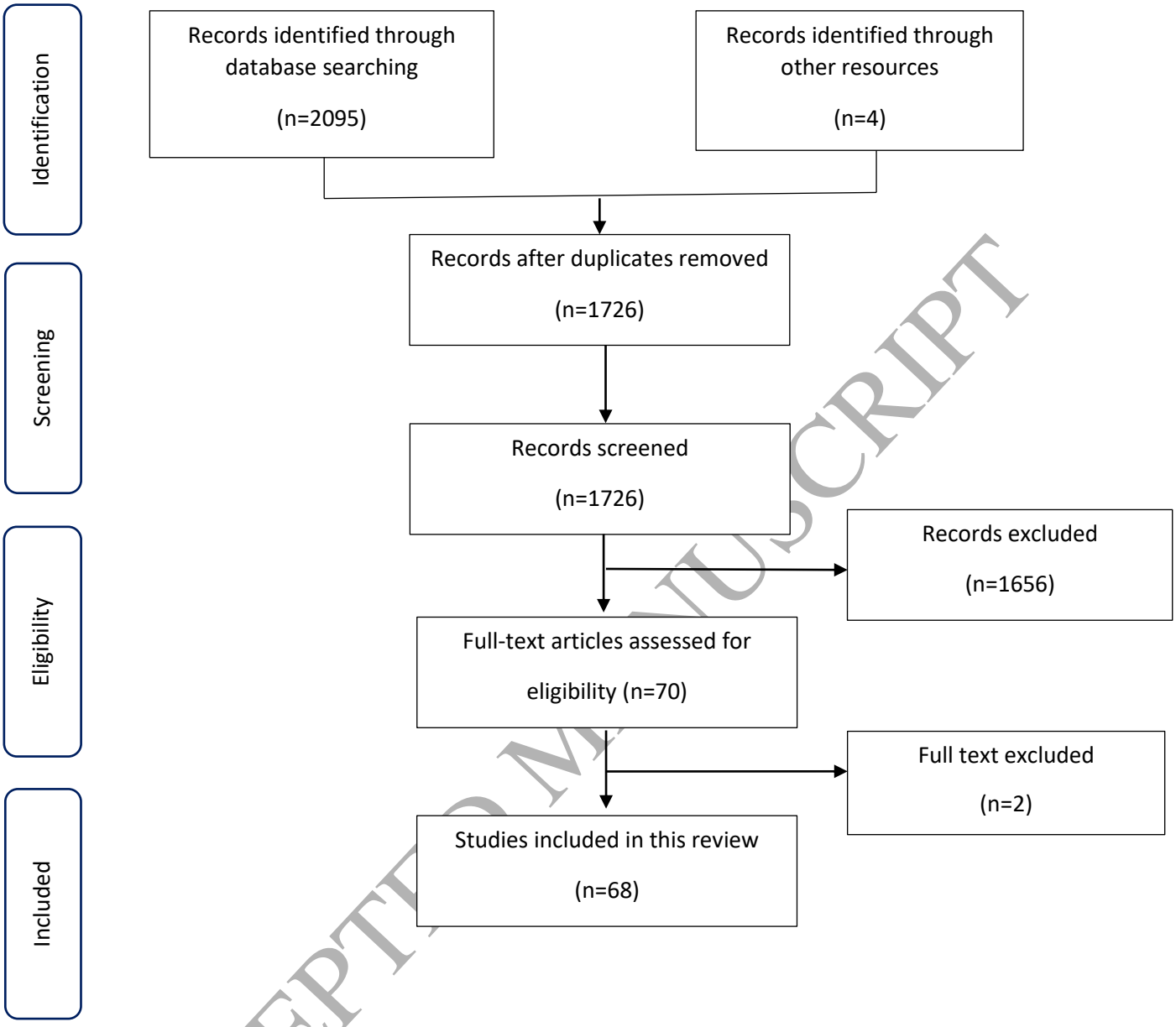

Figure 1: Flowchart of studies selection

Figure 2: The strength of evidence for each instrument.

Right-hand column denotes the percentage of the strength of evidence for all nine COSMIN criteria. Bottom row indicates the percentage of the strength of evidence at the criterion level.

Table 1: Criteria for quality rating of measurement properties

Measurement properties	Criteria for 'adequate' rating
Internal consistency	Cronbach's alpha is ≥ 0.70.
Reliability	Intraclass correlation coefficient for continuous variables or weighted kappa for ordinal variables is ≥ 0.70.
Measurement error	Smallest detectable change (SDC) is less than minimal important change (MIC) or if the MIC is outside the limit of agreement.
Content validity	The items are relevant to the construct being measured.
Structural validity	Factor analysis shows that the instrument items explain more than 50% of the total variance.
Hypothesis testing	Correlation with other instruments that assess child abuse (convergent validity) is ≥ 0.50, or ≥ 75% of the findings are accordance with the *a priori* hypothesis. If the hypothesis tested for a correlation with another construct other than childhood adversity (divergent validity), a score of 'unclear' is given.
Cross-cultural validity	Factor analysis, logistic regression, or item response theory techniques detected differential item function between two or more language groups for the instrument.
Criterion validity	A score of 'unclear' is given for this criterion for all instruments because there is no 'gold standard' for measuring child abuse.
Responsiveness	Correlation with the change score of an instrument assessing child abuse is ≥ 0.5, or if ≥ 75% of the findings were in accordance with a priori defined hypotheses, or if the area under the receiver operator characteristic curve (ROC) ≥ 0.70.
Interpretability	Information on floor or ceiling effects and the minimal important differences are provided.

Table 2: Methodological quality and quality of measurement properties of the instruments

Methodological quality	%	*n*	Instruments
Internal consistency			
Excellent	8%	4	BCAS (+), CAT (+), CEVQ (+), CTQ-SF (+)
Good	20%	10	AEIII (+), CAMI (+), CAQ (+), CCMI (+), CCMS-A (+), CCMS-P (+), CTQ (+), ETI-SR (+), FEQ (+), TEC (+)
Fair	2%	1	FAST (+)
Poor	12%	6	CVS (+), CECA-Q (+), ETI (+), TEC* (+), ICAST-R (+), ICAST-C (+)
Unknown	58%	30	ACE, ADQ, AH, ASE, BPSAQ, CANIS-R, CCMS-P, CECA, CEQ, CHUS, CMHSR CTES, CTI, DI, EHEI, FEI, HPSAQ, LEQ, LTVH, MACE, PMI, RATE, RFPQ, RSEQ SAEQ, SEQev, SEQex, SLEI, SPAQ, STI, TAI
Reliability			
Excellent	0%	0	–
Good	32%	17	ADQ (+), AEIII (+), CAMI (+), CAT (+), CCMI (+), CCMS-A (+), CCMS-P (+), CECA-Q (+), CTI (+), CTQ (+), ETI (+), FEI (+), HPSAQ (+), MACE (+), PMI (+), SEQex (-), TEC(+)
Fair	4%	2	EHEI (+), LEQ (+)
Poor	8%	4	CANIS-R (?), CHUSE (-), RATE (+), SAEQ (+)
Unknown	56%	29	ACE, AH, ASES, BCAS, BPSAQ, CAQ, CECA, CEQ, CEVQ, CMHSR, CTES, CTQ-SF, CVS, DI, ETI-SR, FAST, FEQ, ICAST-C, ICAST-R, LEQ, LTVH, RFPQ, RSEQ, SEQev, SLEI, SPAQ, STI, TAI, TEC*
Measurement error			
Excellent	0%	0	–
Good	2%	1	MACE (+)
Fair	0%	0	–
Poor	0%	0	–
Unknown	98%	51	ACE, ADQ, AEIII, AH, ASES, BCAS, BPSAQ, CAMI, CANIS-R, CAQ, CAT,CCMI, CCMS-A, CCMS-P, CECA, CECA-Q, CEQ, CEVQ, CHUSE,CMHSR, CTES, CTI, CTQ, CTQ-SF, CVS, DI, EHEI, ETI, ETI-SR, FAST, FEI, FEQ, HPSAQ, ICAST-C, ICAST-R, LEQ, LTVH, PMI, RATE, RFPQ, RSEQ, SAEQ, SEQev, SEQex, SLEI, SPAQ, STI, TAI, TEC*, TEC, TEQ
Content validity			
Excellent	23%	12	AEIII (+), CCMI (+), CECA (+), CEQ (+), CEVQ (+), CTQ (+), ETI (+), ETI-SR (+), HPSAQ (+), ICAST-C (+), ICAST-R (+), LTVH (+)
Good	11.5%	6	ADQ (+), DI (+), FAST (+), MACE (+), PMI (+), ACE (+)
Fair	2%	1	SLEI (+)
Poor	0%	0	–
Unknown	63%	33	AH, ASES, BCAS, BPSAQ, CAMI, CANIS-R, CAQ, CAT, CCMS-A, CCMS-P, CECA-Q, CHUSE, CMHSR, CTES, CTI, CTQ-SF, CVS, EHEI, FEI, FEQ, LEQ, RATE, RFPQ, RSEQ, SAEQ, SEQev, SEQex, SPAQ, STI, TAI, TEC*, TEC, TEQ
Structural validity			
Excellent	6%	3	CAT (+), CTQ-SF (+), ETI-SR (+)
Good	2%	1	CTQ (+)
Fair	0%	0	–
Poor	0%	0	–
Unknown	92%	48	ACE, AH, ADQ, AEII, ASES, BCAS, BPSAQ, CAMI, CANIS-R, CAQ, CCMI, CCMS-A, CCMS-P, CECA, CECA-Q, CEQ, CEVQ, CHUSE, CMHSR, CTES, CTI, CVS, DI, EHEI, ETI, FAST, FEI, FEQ, HPSAQ, ICAST-C, ICAST-R, LEQ, LTVH,

			MACE, PMI, RATE, RFPQ, RSEQ, SAEQ, SEQev, SEQex, SLEI, SPAQ, STI, TAI, TEC*, TEC, TEQ
Hypothesis testing			
Excellent	2%	1	MACE (+)
Good	25%	13	ACE(?), ADQ (+), BCAS (+), CAT(?), CEVQ(?), CTES(?), CTI (+), CTQ (+), CTQ-SF (+), FAST (+), FEQ(?), PMI (+), TEQ(?),
Fair	17%	9	AH(?), CHUSE(?), FEI(?), LEQ(?), LTVH (+), RFPQ(?), RSEQ(?), SEQex(?), TAI(?)
Poor	0%	0	–
Unknown	56%	29	AEII, ASES, BPSAQ, CAMI, CANIS-R, CAQ, CCMI, CCMS-A, CCMS-P, CECA, CECA-Q, CEQ, CMHSR, CVS, DI, EHEI, ETI, HPSAQ, ICAST-C, ICAST-R, RATE, SAEQ, SEQev, SEQex, SLEI, SPAQ, STI, TEC*,TEC
Cross-cultural validity			
Excellent	0%	0	–
Good	2%	1	CTQ-SF (+)
Fair	0%	0	–
Poor	4%	2	ICAST-C (+), ICAST-R (+),
Unknown	94%	49	ACE, ADQ, AEII, AH, ASES, BCAS, BPSAQ, CAMI, CANIS-R, CAQ, CAT, CCMI, CCMS-A, CCMS-P, CECA, CECA-Q, CEQ, CEVQ, CHUSE, CMHSR, CTES, CTI, CTQ, CVS, DI, EHEI, ETI, ETI-SR, FAST, FEI, FEQ, HPSAQ, LEQ, LTVH, MACE, PMI, RATE, RFPQ, RSEQ, SAEQ, SEQev, SEQex, SLEI, SPAQ, STI, TAI, TEC, TEC, TEQ
Criterion validity			
Excellent	0%	0	–
Good	0%	0	–
Fair	15%	8	CAMI (?), CCMS-A (?), CECA-Q (?), CEVQ (?), LTVH (?), SPAQ (?), TEC (?), TEC* (?)
Poor	0%	0	–
Unknown	85%	44	ACE, ADQ, AEII, AH, ASES, BCAS, BPSAQ, CANIS-R, CAQ, CAT, CCMI, CCMS-P, CECA, CEQ, CHUSE, CMHSR, CTES, CTI, CTQ, CTQ-SF, CVS, DI, EHEI, ETI, ETI-SR, FAST, FEI, FEQ, HPSAQ, ICAST-C, ICAST-R, LEQ, MACE, PMI, RATE, RFPQ, RSEQ, SAEQ, SEQev, SEQex, SLEI, STI, TAI, TEQ
Responsiveness			
Excellent	0%	0	–
Good	0%	0	–
Fair	0%	0	–
Poor	0%	0	–
Unknown	100%	52	ACE, ADQ, AEII, AH, ASES, BCAS, BPSAQ, CAMI, CANIS-R, CAQ, CAT, CCMI, CCMS-A, CCMS-P, CECA, CECA-Q, CEQ, CEVQ, CHUSE, CMHSR, CTES,CTI, CTQ, CTQ-SF, CVS, DI, EHEI, ETI, ETI-SR, FAST, FEI, FEQ, HPSAQ, ICAST-C, ICAST-R, LEQ, LTVH, MACE,PMI, RATE, RFPQ, RSEQ, SAEQ, SEQev, SEQex, SLEI, SPAQ, STI, TAI, TEC*, TEC, TEQ
Interpretability			
Percentage of missing items	42%	22	ADQ, AEII, ASES, CAMI, CAT, CECA-Q, CECA-Q, CEVQ, CHUSE, CTES, CTI, CTQ, CTQ-SF, CVS, EHEI, LEQ, MACE, PMI, RATE, SAEex, TEC*, TEC
Description of how missing items were handled	4%	2	CAT, MACE
Distribution of the (total) scores	12%	6	CAMI, CCMS-A, CEVQ, MACE, SAEQ, TEQ
Percentage of the respondents who had the highest possible	12%	6	CAMI, CCMS-A, CEVQ, MACE, SAEQ, TEQ

(total) score			
Scores and change scores	12%	6	CAMI, CCMS-A, CEVQ, MACE, SAEQ, TEQ
Minimal Important Change (MIC) or Minimal Important Difference (MID)	0%	0	–

Measurement quality: 'adequate' (+), 'not adequate' (-) or 'unclear' (?)

ABREVIATIONS:
ACE: Adverse Childhood Experiences
ADQ: Anatomical Doll Questionnaire
AEII: Assessing Environments III
AH: Abuse History
ASES: Abusive Sexual Exposure Scale
BCAS: Binghamton Childhood Abuse Screen
BPSAQ: Brief Physical and Sexual Abuse Questionnaire
CAMI: Computer Assisted Maltreatment Inventory
CANIS-R: Child Abuse and Neglect Interview Schedule-Revised
CAQ: Childhood Adversity Questionnaire
CAT: Child Abuse and Trauma Scale
CCMI: Comprehensive Childhood Maltreatment Inventory
CCMS-A: Comprehensive Child Maltreatment Scales for Adults
CCMS-P: Comprehensive Child Maltreatment Scales for Parents
CECA: Childhood Experience of Care and Abuse
CECA-Q: Childhood Experience of Care and Abuse
CEQ: Childhood Experiences Questionnaire
CEVQ: Childhood Experiences of Violence Questionnaire
CHUSE Childhood Unwanted Sexual Events
CMHSR: Child Maltreatment History Self-Report
CTES: Childhood Traumatic Event Scale
CTI: Childhood Trauma Interview
CTQ: Childhood Trauma Questionnaire
CTQ-SF: Childhood Trauma Questionnaire Short Form
CVS: Childhood Violence Scale
DI: Developmental Interview
EHEI: Early Home Environment Interview
ETI: Early Trauma Inventory
ETI-SR: Early Trauma Inventory - Self Report
FAST: Family Aggression Screening Tool
FEI: Family Experiences Interview
FEQ: Family Experiences Questionnaire
HPSAQ: History of Physical and Sexual Abuse Questionnaire
ICAST-C: ISPCAN Child Abuse Screening Tools - Children's version
ICAST-R: ISPCAN Child Abuse Screening Tools Retrospective version
LEQ: Life Experience Questionnaire
LTVH: Lifetime Trauma and Victimization History
MACE: Abuse Chronology of Exposure Scale
PMI: Psychological Maltreatment Inventory
RATE: Retrospective Assessment of Traumatic Experience
RFPQ: Retrospective Family Pathology Questionnaire
RSEQ: Retrospective Separation Experience Questionnaire
SAEQ: Sexual Life Events Inventory
SEQev: Sexual Abuse Questionnaire
SEQex: Sexual Experience Questionnaire
SLEI: Sexual Life Events Inventory
SPAQ: Sexual and Physical Abuse Questionnaire
STI: Structured Trauma Interview,
TAI: Traumatic Antecedents Interview
TEC*: Traumatic Experience Checklist
TEC: Trauma Experiences Checklist
TEQ: Traumatic Experiences Questionnaire